NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

THESIS

INDEXING AND RETRIEVAL IN DIGITAL LIBRARIES: DEVELOPING TAXONOMIES FOR A REPOSITORY OF DECISION TECHNOLOGIES

by

Patricia M. Rogers

March, 1996

Thesis Advisor: Hemant Bhargava

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1996

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
INDEXING AND RETRIEVAL IN DIGITAL LIBRARIES: DEVELOPING TAXONOMIES FOR A REPOSITORY OF DECISION TECHNOLOGIES

5. FUNDING NUMBERS

6. AUTHOR(S) Rogers, Patricia M.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

DecisionNet is an online Internet-based repository of decision technologies. It links remote users with these technologies and provides a directory service to enable search and selection of suitable technologies. The ability to retrieve relevant objects through search mechanisms is basic to any repository's success and usability and depends on effective classification of the decision technologies. This thesis develops classification methods to enable indexing of the DecisionNet repository.

Existing taxonomies for software and other online repositories are examined. Criteria and principles for a good taxonomy are established and systematically applied to develop DecisionNet taxonomies. A database design is developed to store the taxonomies and to classify the technologies in the repository. User interface issues for navigation of a hierarchical classification system are discussed. A user interface for remote World Wide Web users is developed. This user interface is designed for browsing the taxonomy structure and creating search parameters online. Recommendations for the implementation of a repository search mechanism are given.

14. SUBJECT TERMS
1. Decision Support Systems, Digital Repository, Taxonomy, Internet, Models, Algorithms

15. NUMBER OF PAGES 70

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18 298-102

INDEXING AND RETRIEVAL IN DIGITAL LIBRARIES: DEVELOPING TAXONOMIES FOR A REPOSITORY OF DECISION TECHNOLOGIES

Patricia Rogers
Lieutenant , United States Navy
B.A., University of California, Los Angeles , 1991

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
March 1996

Author: Patricia M. Rogers

Patricia M. Rogers

Approved by: HKBhargava

Hemant Bhargava, Thesis Advisor

Gordon Bradley, Associate Advisor

Reuben T Harris

Reuben T. Harris, Chairman
Department of Systems Management

ABSTRACT

DecisionNet is an online Internet-based repository of decision technologies. It links remote users with these technologies and provides a directory service to enable search and selection of suitable technologies. The ability to retrieve relevant objects through search mechanisms is basic to any repository's success and usability and depends on effective classification of the decision technologies. This thesis develops classification methods to enable indexing of the DecisionNet repository.

Existing taxonomies for software and other online repositories are examined. Criteria and principles for a good taxonomy are established and systematically applied to develop DecisionNet taxonomies. A database design is developed to store the taxonomies and to classify the technologies in the repository. User interface issues for navigation of a hierarchical classification system are discussed. A user interface for remote World Wide Web users is developed. This user interface is designed for browsing the taxonomy structure and creating search parameters online. Recommendations for the implementation of a repository search mechanism are given.

TABLE OF CONTENTS

LIST OF ACRONYMS

AT&T- American Telephone and Telegraph, Inc.

CGI - Common Gateway Interface

COE- Common Operating Environment

CPU- Central Processing Unit

DoD - Department of Defense

DSS- Decision Support System

DT - Decision Technology

GAMS - Generalized Algebraic Modeling System

GAMS - Guide to Available Mathematical Software

GAO - Government Accounting Office

HTML- HyperText Markup Language

IBM - International Business Machines, Inc.

NAG - Numerical Algorithms Group

NIST - National Institute of Standards and Technology

RAD - Rapid Application Development

SP - Search Parameters

SQL - Structured Query Language

TCP/IP - Transmission Control Protocol /Internet Protocol

WATERSHEDSS - Water, Soil, Hydro-Environmental Decision Support System

WAIS - Wide Area Information Service

WWW - World Wide Web

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

The widespread popularity and growth of the Internet and the World Wide Web has created new opportunities for the sharing of data and executable software between users worldwide. One such application is DecisionNet, an online repository of decision support systems. The general area of research this thesis addresses is the implementation issues involved with establishing an online repository. The repository is distributed (objects are stored remotely on other servers worldwide) but cataloged at the Naval Postgraduate School and accessed centrally by users worldwide through DecisionNet. The DecisionNet system consists of the decision technologies, a database of information describing those objects and software agents that handle consumer and provider administrative information and perform browsing and search mechanisms. (Bhargava, King and McQuay, 1995)

The objective of this research is to develop and implement taxonomies for classifying the technologies for the purpose of indexing and retrieval by providers and users of the system. An effective classification scheme will allow users to search and get results scored by relevancy, and will be useful in doing automatic registration of objects by remote providers.

A. DECISION SUPPORT SYSTEMS

Decision Support Systems (DSS's) are used by managers and planners to perform modeling, simulation and expert analysis of variables and conditions in support of decision making. DSS's are software programs that employ computational methods such as mathematical algorithms and artificial intelligence techniques to determine the best action to take using available information. DSS's are valuable for managers and decision makers in fields ranging from evaluating investment opportunities for personal use to wargaming and logistics modeling for the Department of Defense (DoD). A number of decision support systems apply results from research in Operations Research (OR)/Management Sciences (MS) using techniques such as optimization and modeling.

Some examples of optimized results given resources under constraint are:

- Scheduling and assignment of police officers
- Reducing transportation time and fuel cost
- Reducing fuel cost to DoD for transport of supply/personnel
- Inventory management
- Determining risk tolerance and investment selection
- Production planning
- Selection of locations for service providers, utility stations, refueling stations

These examples are illustrative of civil government functions, military, commercial and private uses of DSS's and their use in improving and analyzing decisions. Potential users of DSS's and DecisionNet include university researchers and students, military and government planners, private industry staff and other Internet users.

B. DECISIONNET: WHY NETWORK BASED DECISION SUPPORT SYSTEMS?

The Internet has become an important medium for the interchange of ideas, data, software and other unique problem solving applications. A DSS repository such as DecisionNet will link users, operations research problem solvers and information system developers.

DSS's tend to be narrowly focused on a specific problem or industry (Sprague, 1980) and while they are effective tools for decision makers they may be costly to develop or purchase. The ability to share software on a network such as the Internet means software is made available for use without necessarily requiring purchase. Software is essentially rented and intellectual property rights or solution methods remain in the hands of the developer (Bhargava, King and McQuay, 1995).

Using the Internet, data is sent to the objects (models, algorithms, and decision support systems) or responded to online through the use of forms and browsable World Wide Web pages in a client-server configuration. For some DSS's, data is sent by the user to the remote decision technology where the answer is computed and sent back to the user. For other DSS's, there is an interactive text based dialogue between the client and server. For example, the Water, Soil, Hydro-Environmental Decision Support System (WATERSHEDSS) is a knowledge based water quality decision support system. To use the DSS the user navigates a decision tree to arrive at a text based answer. For both types, the DecisionNet system provides the directory service and "mediates the initial transactions between users and different technologies." (McQuay, 1995)

The Internet is based on the Transmission Control Protocol/Internet Protocol (TCP/IP) as a standard for transmitting all kinds of files, including executable software, e-mail, data sets and World Wide Web pages. These tools and infrastructure make the Internet the perfect platform for this type of application.

The DoD has also adopted the TCP/IP standards for the common operating environment (COE) systems of the future. By developing DecisionNet within the standards of these Command and Control Systems, a parallel system could be developed to function operationally on military networks, similar to the Internet. Such a system might primarily store military DSS's for military decision makers. (Defense Information Systems Agency, 1994)

C. DECISIONNET AS A DIGITAL LIBRARY

DecisionNet is designed to be a digital library for DSS technologies. In order to keep track of the objects in the DecisionNet repository, the meta-information about objects must be organized logically to facilitate searching of the database by users.

Digital libraries share important characteristics with physical libraries in use today. Libraries are repositories for many types of objects. In order to organize these objects for quick and meaningful retrieval (based on a search by a user,) they are described and classified by their characteristics. The library's collection, like DecisionNet's repository, is represented by a database and the user interacts with a system that accesses the database.

Physical libraries generally keep many types of objects: newspapers, magazines, films, videotapes, books, pamphlets, etc. . The characteristics on which objects are classified include:

- Each object has a primary author, editor or creator.
- Each object has a publisher or distributor and date of creation or publication.
- Each object has a title.
- Each object is classified by subject. (Libraries use taxonomies like a subject tree or the Dewey Decimal System for non-fiction works.)

These characteristics or properties together describe one unique object. Even if the only difference between one object and another is edition number (different date of publication with some modifications and improvements,) no two objects are identical.

As in a physical library, the contents of a digital library are identified by their unique characteristics. The user has to be introduced to a subject tree or taxonomy that he navigates to locate the object he is seeking.

We have determined the unique characteristics of objects contained within DecisionNet to be:

- *Problem area-* the specific task performed by the DSS
  e.g. loan amortization, water quality determination

- *Functional area - the specific* departments within an organization
  e.g. administration, operations, legal, sales

- *Industry type -* the type of work a business or organization performs
  e.g. engineering, manufacturing

- *Organization Type-* one of the five types of user orientations for which the application was designed: military, government, non-profit, commercial, or individual

- *Object type-* one of the four types of objects DecisionNet stores: data set, algorithm, model, or functional decision support system

- *Solution Method-* the mathematical or deductive method of the DSS
  e.g. optimization, numerical math, spreadsheet modeling

Some users of physical libraries go in with an idea about at least one of the characteristics they are looking for and some knowledge of the subject area and browse (a computer or a shelf of the library) until they find the best one or two objects. Other users know exactly what they want and only follow the signs in the library pointing them to its location. First time users need to start with a tour of all available services.

Similarly, digital library users will search for topics in different ways. A casual Internet user looking for a simple DSS like a mortgage calculator may use the system and have no knowledge or interest about how such a task is performed. Another user, perhaps an undergraduate student taking a computer course, may need to find out about decision support systems but not have any real idea about the solution method they seek. Some users will just want to browse the repository to see what is available. Operations research and computer science researchers may seek a specific computational solution method regardless of the application it performs.

This thesis explores the problem of how to organize and classify the objects in DecisionNet repository based on their unique characteristics: *problem area, functional area, industry, organization type, object type,* and *solution method.* Effective classification of the elements of the repository will allow users to browse the repository for what they are looking for by themselves and allow for searches based on the characteristics of the object they seek.

D. PREVIOUS WORK DONE ON DECISIONNET

Prior to this thesis, work on DecisionNet identified the required primary capabilities and information flow. The first prototype automated many of the registration functions and built a search mechanism for registered DSS technologies by keyword search of flat files containing all information available about the objects. Both CAPT Dan McQuay (McQuay, 1995) and LT Andy King (King, 1995), who developed the first prototype, highlighted the need for a taxonometric classification scheme to implement relevancy searches.

E. DEFINITION OF KEY TERMS AND CONCEPTS

1. DecisionNet System Components

The DecisionNet system is comprised of three basic parts: informational HTML pages and software agents, a database, and the repository. The repository is the collection of decision technologies that are registered in DecisionNet. The software agents handle the user's input through interaction with the database tables. The database contains information about the decision technologies that the search mechanisms query to match the user's request. The system directs the user to the objects in the repository that match the search request. Users of the system are classified as consumers or providers. Consumers search the database looking for a decision technology to use and providers register technologies others might want to use.

2. Taxonomy Definitions

A taxonomy is an abstract division of objects into ordered groups or categories based on their characteristics. One of the most familiar taxonomies is used in biology to divide all organisms into one of two kingdoms: animal and plant. Within a kingdom, objects are broken down into phyla, then classes, orders, etc. as shown in Fig. 1.1.

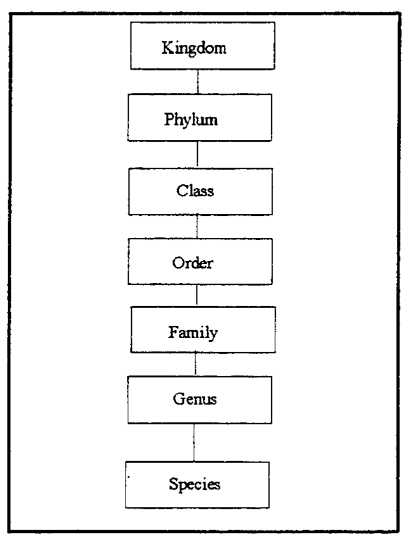

Figure 1.1 Taxonomy Applied to Organisms

Each level is considered a "taxon", a taxonomic category or group. To list family types and genus types together would be to confuse levels of specialization. Taxonomies are similar to family trees because properties of inheritance exist. Organisms within the same order have similar characteristics in common but belong to different geni or species because of their distinct characteristics.

3. Object Types Handled by DecisionNet

The DecisionNet repository catalogs four types of objects: data sets, algorithms, models and functional decision support systems.

F. SEARCH TYPES

In assessing how searches of the database should be processed there are two primary considerations, how the search is executed and how the database is indexed. One search method is a batch or real time search of the database that involves searching the database on each request. The results are always current, and this method is appropriate for a dynamically changing database. Another search method is to index the database and make updates on a weekly or daily basis. This is a faster approach, but not necessarily completely current or perfectly accurate. Current practice on the Internet is that repositories that have a database interface do batch processing of requests. Online Internet search engines primarily do queries to a central database that is periodically updated. These updates are made by traveling spiders or indexing robots. The appropriate search method depends on the objective of the administering organization. Is it to minimize retrieval time? Maximize accuracy and thoroughness? Are there serious constraints on the storage space and processing power of the database? Each of these leads to a different architecture of the database.

Another consideration is that a database can be indexed for searching in two ways. The first method is keyword searching the entire database for a certain string or term. For some keyword searches there will be many exact matches. Because synonyms or a similar concept may be used to classify the object under a keyword search orientation there is a chance nothing would be retrieved and that relevant objects would not be brought to the attention of the user. The second way is to compute relevancy based on classification of objects under an indexing scheme. In order to include all relevant items and to allow users to find what they are looking for, a relevancy calculating scheme is used. To compute relevancy a classification scheme for each of the six classifying characteristic must be developed. Ideally every conceivable choice under each of these characteristics would be identified to build a mutually exclusive and collectively exhaustive list of selections. To take this organization one step further, we want to teach the search mechanisms the taxonomy of each classifying characteristic so that it will retrieve something related if the exact query can not be satisfied.

G. ORGANIZATION OF THESIS

This thesis is organized into five chapters and one appendix. The first chapter contains the introduction, an overview of the DecisionNet concept, a discussion of the work done on system design, and a discussion of search mechanisms for a repository. The second chapter examines the relationship between the objects in the repository and their representation in the database. Additionally, the concept of taxonomies is introduced as the primary indexing scheme, an analysis of current practice is done and the criteria for a good taxonomy are developed. Chapter III discusses the methodology used to develop taxonomies for four of the six characteristics. Appendix A is the expanded view of the solution method taxonomy that is shown in Chapter III. Chapter IV examines the implementation of the taxonomies using available Internet software tools. That chapter also discusses principles of user interface design considered in developing the DecisionNet system's user interface. Chapter V is the thesis conclusion and offers suggestions for future work on the DecisionNet classification scheme and taxonomy development.

II. TAXONOMIES: ROLE & CRITERIA

A. SEARCHING A DATABASE - HOW TO ORGANIZE A SITE FOR OUTSIDE USERS

The current practice for Internet online pages is to include text, graphics and logos, product information and personnel and point of contact information. Increasingly organizations are seeing the Internet as a forum for network publishing. For example, government publications are increasingly available online because this is an inexpensive way to publish them. Network publishing provides access by a greater number of users who were previously limited by physical proximity to a database or libraries holdings. Examples include:

- Digital Libraries- Library of Congress and university libraries worldwide
- Bookstores/ video and music stores that make some of their products available for worldwide users
- Government Printing Office- legislation, Census Information, Government Accounting Office (GAO) reports

These repositories have the common need to classify their collections for indexing and retrieval by remote users. The Wide Area Information Service (WAIS) is a software package that indexes the entire contents of a server and has an online search engine. Results are scored on the number of occurances of a word (keywords entered by the user) within the documents that meet certain search parameters (e.g. written after 1993) . For example, the Government Accounting Office (GAO) has put all its reports online and they are completely indexed and available for search and retrieval by keyword using WAIS. This is effective because the objects are homogenous and keyword search effectively distinguishes the primary differences between the objects in the repository. But it does not solve the synonym problem mentioned earlier. More importantly, this kind of indexing scheme is not appropriate for DecisionNet's consideration of more than one characteristic by the search mechanism.

B. USING TAXONOMIES TO CLASSIFY AND SEARCH FOR OBJECTS

To give a preview of the DecisionNet taxonometric classification scheme, there are six characteristics every object has and can be described by (the horizontal axis in the following figure). Under each characteristic is a hierarchical taxonomy (represented vertically). The depth of the classification represents how many levels of specialization exist within that type. The more classification and specialization assigned to an object the greater the capability for finer and more accurate searching.

Figure 2.1 illustrates that decision technologies possess characteristics and that a relationship exists between those objects and the taxonomy of each characteristic.

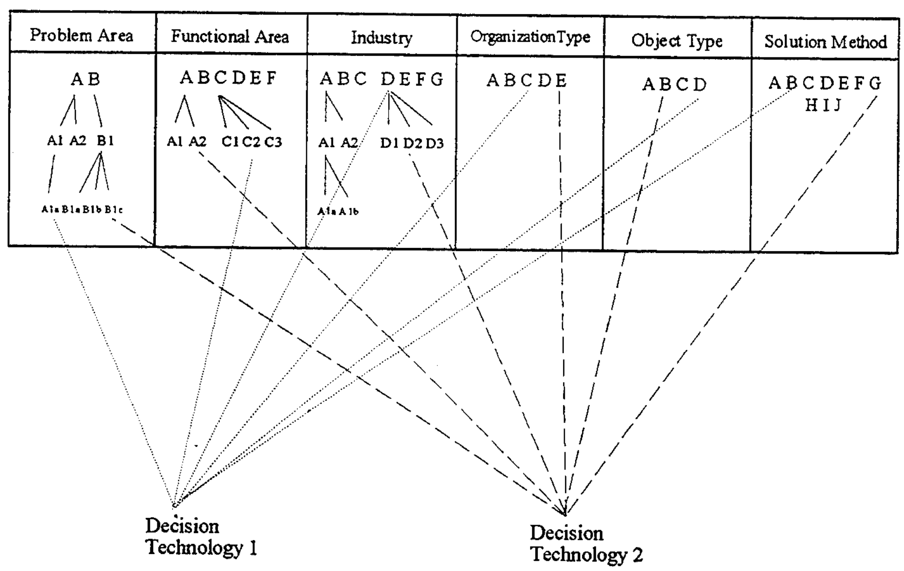

Figure 2.1 Relationship Between Decision Technology Characteristics and Taxonomy Scheme

Essentially, the characteristics exist as abstract classes whether or not there are any objects with those characteristics currently registered in the repository. In Figs. 2.1 and 2.2 Decision Technology 1 has the following attributes: problem area (A1a), functional area (C2), industry (D), organization type (C), object type (D), and solution method (B). Decision Technology 2 has the following attributes: problem area (B1c), functional area (A2), industry (D), organization type (E), object type (B) and solution method (G). In Fig. 2.2 the user's search parameters are scored against the decision technologies in the repository (as they are represented in the database). The search mechanism computes relevancy by determining the distance between nodes representing the users search parameters and the decision technologies in the repository.

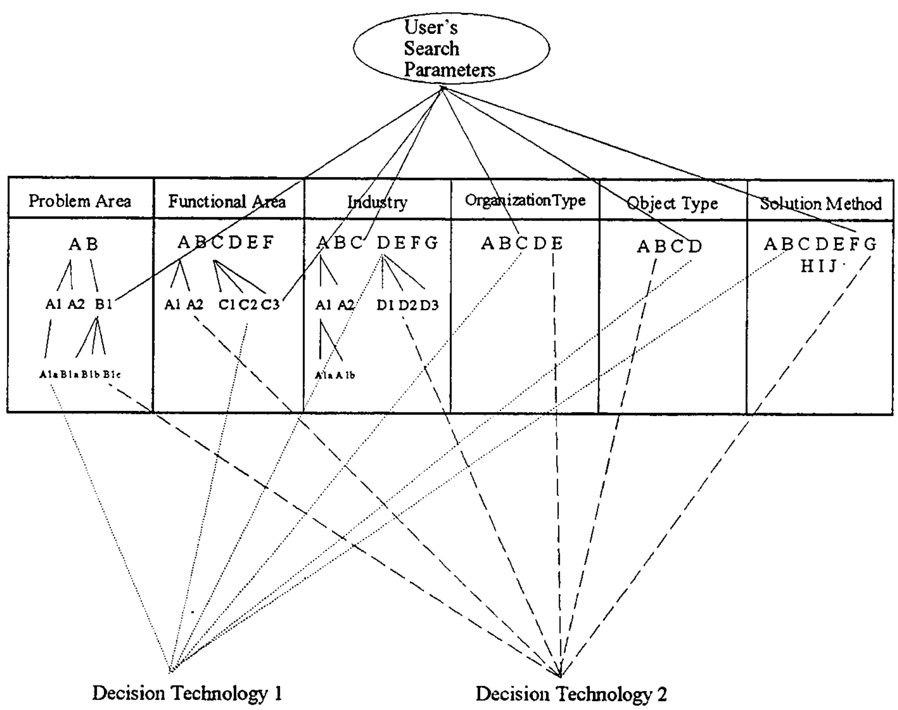

Figure 2.2 Searching DecisionNet Repository by Characteristic

Using Decision Technology 1 and 2 for this example, a simple search mechanism compares the user's request against the characteristics of the two objects and returns a rating as to which is more relevant. Relevancy is determined by computing the distance between a specific decision technology (DT) and user's search parameters (SP) to a common parent in each characteristic.

	User SP	DT 1	DT 2	Distance User's SP to DT1	Distance User's SP to DT2
Problem Area	B1	A1a	B1c	3	1
Functional Area	C3	C2	A2	1	3
Industry	C	D	D	1	0
Organization Type	C	C	E	0	1
Object Type	D	D	B	0	1
Solution Method	F	B	G	1	1
Total Distance				5 nodes	7 nodes

Table 2.1 Distance Between Decision Technologies and User's Search Parameters

Decision Technology 1 is more relevant to the users search than Decision Technology 2 and would be scored accordingly.

C. LITERATURE REVIEW - CURRENT PRACTICE

One of the most important contributions of the Internet is the easy access it provides to documents, algorithms, software modules, data sets and executable programs. A number of repositories have been established by academic and commercial organizations and these repositories face the organization of information and search problems discussed previously. There is little information published that is directly related to implementing a network based repository.

There are many keyword search engines on the Internet and very few classification schemes. Two repositories that do effectively index their collections with classification schemes are Yahoo and the Guide to Available Mathematical Software (GAMS). The first, Yahoo, is a "searchable, browsable, hierarchical index of the Internet" by subject. The second, GAMS, is an index of algorithms by solution method. The primary contribution of GAMS is its problem classification system. The GAMS, developed by the National Institute of Standards and Technology (NIST), is a 736 node "tree structured taxonomy extending to seven levels. Each child node in the tree denotes a more specific instance of the mathematic problem represented by its parent". (Boisvert, 1991)

Many academic and some commercial organizations are developing online repositories of software or algorithms that are organized using keyword searches or adopting the GAMS problem classification scheme. The Numerical Algorithm Group (NAG) and Netlib (developed in conjunction with Oak Ridge National Laboratory and the AT&T Bell Labs) both use GAMS as the single classification characteristic for their repositories. The DecisionNet repository, however, will keep more than just algorithms. It will also classify models and DSS's that are tailored to a specific application and user group.

D. TAXONOMIES WITHIN A REPOSITORY

A taxonomy is defined as the division into ordered groups or categories. Systematics or a taxonomic organization of a field is undertaken to "structure the body of knowledge that constitutes a field once it has reached a certain level of maturity" (Glass and Vessey, 1994). This is done to break a field down for further study, to understand the elements of a field, and to predict future areas of study or to create specialization.

Taxonomies are essentially hierarchies where objects are classified from general to specific. Once we have decided the dimensions that describe the objects, what will the taxonomy of options look like? A tree is the most natural for a classification scheme because it "allows for refinement in mature and young subject areas". Hierarchical based classification gives the users the flexibility to refine the specification by "using the classification as a decision tree" (Boisvert, Aug. 1994) and allows a way for matters of relevance and relatedness to be determined so that they can be built into the search routine. Classes will be linked and may be "siblings" or "children" of other areas. Selections should also be mutually exclusive and collectively exhaustive.

E. CURRENT PRACTICE FOR DSS TAXONOMIES

Consider current practice and academic consideration of DSS taxonomies. A literature review revealed that the Decision Support Systems community of the Information Systems field and the software development side of the OR/MS field think of decision support systems in broad terms. Table 2.1 illustrates one method for classifying DSS's:

Type of problem	Generic Approach	Nature of Choice
Strategic focus	Analyzing	Optimal
Tactical focus	Standardizing	Acceptable
Operational focus	Automatizing	Rewarding

Table 2.2 (After Nagel, 1993) Decision Support System Problem Types

This view considers structured (operational), semi-structured (tactical) and unstructured (strategic) decisions as an evaluation between degree of risk and risk aversion, payoff or loss potential and certainty of information.

The Glass and Vessey paper "Towards Taxonomies of Application Domains"(1994) considers several taxonomies in use by the DoD, software development companies and other software standards making bodies. Their analysis of these taxonomies is that software was broken down as either application oriented or infrastructure oriented and within those domains it was considered to be either problem focused (industry or application - called problem area in this paper) or solution focused (called solution method in this paper). Another DSS researcher categorized decision support systems as:

- Approaches based on OR/MS
- Spreadsheet based
- Expert system based (rule based/function based) (Nagel, 1993)

While these may be effective for considering a single DSS or comparing one with another it is not an effective way to describe and search a repository of DSS's. However, Nagel's generalization was a starting point for creating a taxonomy based on solution method as discussed later.

Each of these examples is not suitable for classifying a decision support system in a way that helps the user find what they're looking for in a repository of many objects.

F. CRITERIA FOR TAXONOMIES

Dr. Boisvert, the creator of GAMS writes "To be effective a classification system must have the following properties:

- problem-orientation
- variable-level tree structure for specialization and generalization of topics
- active maintenance by system administrators for refinement" (1991)

Based on these classifications and other examples it was decided that the following criteria should be used: a taxonomy be principled, scalable, maintainable and robust. Principled, means that it consists of elements that are distinct instances of the same attribute at a given level and those distinctions are based on the principles that apply to that characteristic. Maintainable refers to the capacity of each taxon to be thorough and create significant distinction between the categories while not being stifling at the highest level. A scalable taxonomy is one that although it is thorough at each level the structure is built in such a way that further refinement or specialization can be done at the lower levels without changing the original structure. Robust refers to the completeness of the original framework in describing the problems the repository will be applied to so that the original architecture covers the fields and data sorting and classifying needs of the system in its future.

III. DEVELOPMENT OF TAXONOMIES FOR DECISION TECHNOLOGIES

The primary objective of this thesis is to identify the classifying characteristics or meta-information required for each decision technology in the database and then develop a taxonomy for each characteristic. Providers use the taxonomy to register objects and consumers use them to search the database.

The six ways in which objects can be classified or searched are:

- problem area
- functional area
- industry
- organization type
- object type
- solution method

The object types to be handled by DecisionNet was settled on previously and the organization type was easily classified into five types and did not require further classification, but the other four required taxonomy development.

For the task of creating a taxonomy for each characteristic there is a great deal of material on existing indexing methods. Some of the materials available includes the subject classification scheme used in the Operations Research/ Management Science Index (Tolle, 1988), the Encyclopedia of Operations Research and Management Science (Gass and Harris, 1995) and the Guide to Available Mathematics Software (GAMS). Table 3.1 shows the GAMS framework at the highest level with amplification of the optimization class to three additional levels.

To build a taxonomy for the other characteristics existing classification frameworks were studied. For *problem area, functional area* and *industry* we began with types of businesses and services and their departmental components and the specific tasks they performed. The OR/MS community has done this in an effort to classify the articles written by members of their community and developed a primary subject classification system used in the OR/MS indices. However the subject areas are not of the same taxon, that is to say, at the same taxonomic level or within the same classification characteristic. However it provided a valuable starting place for topics and industries involved with Decision Support Systems.

All fields and disciplines regardless of whether they may or may not have any current association with decision support systems were considered. All problem areas or functions that potentially could be analyzed or modeled with the aid of a decision support system were included. This was to keep the taxonomies robust.

Glass and Vessey's paper provided a number of taxonomies developed by industry software leaders broken down by application type and their component tasks. The paper argues for the importance of "applying strong application-dependent methods in the computing field" particularly in expert systems and knowledge based systems (Glass and Vessey, 1994). Additionally, the U.S. Census and U.S. Patent Office list very specific business types, functions and methods of industries.

A. Arithmetic, error analysis
B. Number theory
C. Elementary and special functions
D. Linear Algebra
E. Interpolation
F. Solution of nonlinear equations
G. Optimization
    G1. Unconstrained
    G1a. Univariate
    G1a1. Smooth function
    G1a2. General function (no smoothness assumed)
    G1b. Multivariate
    G1b1. Smooth function

G1b2. General function (no smoothness assumed)
G2. Constrained
G2a. Linear programming
G2a1. Dense matrix of constraints
G2a2. Sparse matrix of constraints
G2b. Transportation and assignments problem
G2c. Integer programming
G2c1. Zero/one
G2c2. Covering and packing problems
G2c3. Knapsack problems
G2c4. Matching problems
G2c5. Routing, scheduling, location problems
G2c6. Pure integer programming
G2c7. Mixed integer programming
G2d. Network (for network reliability search class M)
G2d1. Shortest path
G2d2. Minimum spanning tree
G2d3. Maximum flow
G2d4. Test problem generation
G2e. Quadratic programming
G2e1. Positive definite Hessian (i.e., convex problem)
G2e2. Indefinite Hessian
G2f. Geometric programming
G2g. Dynamic programming
G2h. General nonlinear programming
G2h1. Simple bounds
G2h2. Linear equality or inequality constraints
G2h3. Nonlinear constraints
G2i. Global solution to nonconvex problems
G3. Optimal control
G4. Service routines
G4a. Problem input (e.g., matrix generation)
G4b. Problem scaling
G4c. Check user-supplied derivatives
G4d. Find feasible point
G4e. Check for redundancy
G4f. Other
H. Differentiation, integration
I. Differential and integral equations
J. Integral transforms
K. Approximation
L. Statistics, probability

M. Simulation, stochastic modeling
N. Data handling
O. Symbolic computation
P. Computational geometry
Q. Graphics
R. Service routines
S. Software development tools
Z. Other

Table 3.1 GAMS Classification System at Highest Level and all of Optimization Taxon

A. PROBLEM AREA

This taxonomy lists the specific task on which the decision maker would like support from a DSS. This list is created primarily from the OR/MS Index, Encyclopedia of OR/MS , and from examples of DSS's online or developed by students for small scale implementation. One of the most difficult classification decisions was how to list the OR problems that have nicknames (knapsack problem, bin packing) which have one distinct solution method. These were represented under the optimization hierarchy in the solution methods and also listed under problem area where applicable. The final taxonomy of Problem Area or specific task is shown in Table 3.2.

A. Asset pricing
B. Assignment
C. Bin packing
D. Capital budgeting
E. Communications networks
F. Corporate strategy
G. Costs analysis
H. Crew scheduling
I. Depreciation
J. Environment/Agriculture systems analysis
K. Facilities/equipment planning
L. Fire models
M. Hierarchical production planning
N. Inventory
O. Investment

P.	Knapsack problem
Q.	Job shop scheduling
R.	Layout/Location of equipment
S.	Learning/Training
T.	Lifecycle
U.	Location analysis
V.	Maintenance/Repair
W.	Manpower planning
X.	Material handling
Y.	Planning
AA.	Production
AB.	Quality control
AC.	Queuing
AD.	Reliability of systems
AE.	Risk assessment/management
AF.	Safety
AG.	Scheduling/sequencing
AH.	Taxation
AI.	Traffic analysis
AJ.	Utility/Preferences
AK.	Vehicle routing
AL.	Yield management

Table 3.2 Problem Area Taxonomy

B. FUNCTIONAL AREA

Functional area is defined as the component parts of an organization. The characteristic "functional area" was created to bridge the gap between industry and problem area, since tasks are not elements of Industries as much as they are elements of departments or functional areas within Industries. Also this provides a way for users to search objects primarily by their industry and functional area and to a lesser degree by their specific problem area, to see what is available. Conversely, the user could search by problem area and functional area ignoring industry. The functional area listing comes from departments of industry, government and military organizations as suggested by Yahoo and the OR/MS Index. The functional area listing is shown in Table 3.3.

A. Administration
B. Engineering
C. Finance
D. Legal
E. Logistics
F. Maintenance
G. Marketing
H. Payroll
I. Personnel
J. Public Relations
K. Sales
L. Software Delivery
M. Supply
N. Testing
O. Training
P. Transportation
Q. Telecommunications

Table 3.3 Functional Area Taxonomy

C. INDUSTRY

The industry listing defines businesses and organizations by the type of work they do or branch of a field (such as manufacturing). The references in this list are adapted from the Census Bureau, Yahoo, the OR/MS Index, OR/MS Encyclopedia, Reifer's Application Taxonomy (Reifer Consultants, 1990), Digital Corporation (Digital Corp., 1991) and IBM's Industry Taxonomies (IBM Corp., 1988).

- A. Aerospace/Space
- B. Accounting
- C. Agriculture
  - C1. Crops
  - C2. Food Production
  - C3. Livestock
  - C4. Weather
- D. Arts
  - D1. Fine Arts
  - D2. Music
  - D3. Theater
- E. Business/Commerce
  - E1. Apparel
  - E2. General Retailing
  - E3. Mailorder
  - E4. Restaurant
  - E5. Wholesale
- F. Communications
- G. Construction
  - G1. Architecture
  - G2. Materials
- H. Economics
  - H1. Macroeconomics
  - H2. Microeconomics
- I. Education
  - I1. K-12
  - I2. Undergraduate
  - I3. Graduate
  - I4. Career training
  - I5. Vocational
- J. Emergency Services
  - J1. Fire
  - J2. Police

- J3. Hospitals

K. Energy

L. Engineering

M. Entertainment/Media
- M1. Broadcasting
- M2. Film
- M3. Music
- M4. Publishing
- M5. Radio
- M6. Television

N. Environment and Ecology
- N1. Air quality
- N2. Forestry
- N3. Pollution
- N4. Recycling
- N5. Pollution
- N6. Water quality

O. Finance
- O1. Banking
- O2. Insurance
- O3. Investment
- O4. Mortgage
- O5. Personal
- O6. Real Estate

P. Government
- P1. Elections
- P2. Energy Policy
- P3. Public Works
- P3. Regulations
- P4. Tax Policy
- P5. Urban Systems

Q. Health /Medicine
- Q1. Exercise
- Q2. Health care
- Q3. Nutrition

R. International Trade

S. Labor

T. Law

U. Libraries
- U1. Archives

V. Manufacturing
- V1. Aerospace

- V2. Automotive
- V3. Building materials
- V4. Chemical
- V5. Clothing/Textiles
- V6. Forest/ Paper products
- V7. Furniture
- V8. Metals
- V9. Pharmaceuticals
- V10. Plastics

W. Marketing

X. Military
- X1. Avionics
- X2. Command and Control
- X3. Data processing
- X4. Simulation
- X5. Software tools
- X6. Telecommunications
- X7. Testing

Y. Natural Resources

Z. Petro-Chemical

AA. Population

AB. Recreation and Sports

AC. Scientific Research
- AC1. Chemistry
- AC2. Geography
- AC3. Engineering
- AC4. Mathematics
- AC5. Meteorology
- AC6. Oceanography
- AC7. Physics
- AC8. Statistics

AD. Transportation
- AD1. Airline
- AD2. Automobile
- AD3. Cargo
- AD4. Railroad
- AD5. Shipping

Table 3.4 Industry Taxonomy

D. ORGANIZATION TYPE

This characteristic defines the kind of organization for which the object was originally designed. An organization is defined as an association of people brought together for a particular purpose. The organization determines how objectives, risk and utility are weighed; these are key elements in decision support systems and decision analysis. A government or military DSS will not express itself in terms of profitability but by service or "readiness" objective functions. There may be overlaps between the different organization types and in many instances this will not be an important element but to other objects it may be vital to the correct use of an object. Table 3.5 lists the organization types. There is no further taxonomy developed for this characteristic, but specialization could be created as needed.

A. Military
B. Government (non-military)
C. Commercial
D. Non-commercial/non-profit
E. Personal/Individual

Table 3.5 Organization Type Taxonomy

E. SOLUTION METHOD

The solution method taxonomy is primarily a consolidation of parts of the GAMS software repository classification system into six areas with the inclusion of the knowledge based, spreadsheet modeling and decision analysis techniques. Table 3.6 shows the top level of the solution method taxonomy. The full taxonomy to four levels deep is available in the Appendix.

A. Simulation/ stochastic modeling
B. Statistics/probability - problem has elements of uncertainty
C. Gaming
D. Optimization
E. Numerical Math -(non optimization)/(non statistics)
F. Spreadsheet modeling (does not include imbedded optimization) formula/expression evaluation
G. Knowledge based
H. Symbolic math (calculus)
I. Decision Analysis
J. Other

Table 3.6 Solution Method Taxonomy

IV. IMPLEMENTATION OF THE TAXONOMY

To implement the taxonomy classification scheme and make it available to the users it had to become part of the system's user interface. The user interface was built using hypertext markup language (HTML) and common gateway interface (CGI), a data handling technique, with Netscape 2.0 as the de facto browser standard. The available HTML tools and those being developed provide the primary structure for building the User Interface. Primary HTML tools used were forms, drop down menus (vital to browsing and selecting taxonomies,) and "frames." In fact, the proliferation of new browsers and capabilities by different companies posed another design question. For which browsers should the system be built? -low end (lowest common denominator- most users) or high end (greatest capabilities, more tools- but more maintenance involved in keeping the User Interface current and possibly leaving out users who don't keep their browser up to date.) The decision was to build for the high end and greatest capabilities. This is because the implementation of a sophisticated database, software agents and CGI functions require the tools available at the higher end of the spectrum of Internet browsers. Also we anticipate our primary users will use the higher end Internet browsers.

A. THE DECISION TO ORGANIZE DECISIONNET AS A DATABASE SYSTEM

The DecisionNet system was first built on a Unix system using Perl (Wall and Schwartz, 1991) scripts. The database was built using flat file searches implemented with a Harvest indexing system (Bowman et al., 1994). The fundamental difficulty was the inability to do relevancy searches that could access a taxonomy structure using that operating system.

In analyzing the relationships between the users, providers, and objects, and the need to perform relevancy based searches we determined the storage of the object information would be most appropriately handled by a relational database. Delphi, a rapid application development (RAD) tool, with a Paradox based relational database is used.

Delphi uses reusable Pascal scripts and many CGI components that can direct data transferring from remote client user to server. (Calvert, 1995) CGI components are quickly being developed by the Internet users community.

Based on the equipment and infrastructure that is currently available we are:

1. creating a web based interface (HTML pages) using dimensions and taxonomies that users will have to choose from for their search and registration of objects.
2. linking the user interface to the server applications (administrative functions and using registered technologies, probably using Delphi and SQL queries).
3. developing and activating search mechanisms (relevance and retrieval algorithms) that can return relevant scores on all dimensions of objects.

The entire DecisionNet database consists of several tables to do administrative tasks and handle security functions. The basic tables consist of:

- user information
- provider information
- object information
- six master tables

Tables 4.1 and 4.2 illustrate how the DecisionNet system searches the database fields by counting the distance between nodes to a common parent.

Name	Parent
Industry	(None)
Government	Industry
Manufacturing	Industry
Public Works	Government
Tax Policy	Government
Automotive	Manufacturing
Building Materials	Manufacturing
Plastics	Manufacturing

Table 4.1 Industry Master Table

Name	Parent
Solution Method	(None)
Knowledge Based	Solution Method
Numerical Math	Solution Method
Optimization	Solution Method
Decision Analysis	Solution Method
Decision Trees	Decision Analysis
Unconstrained	Optimization
Constrained	Optimization
Linear Programming	Constrained

Table 4.2 Solution Method Table

Each "master" table lists the taxonomy level and its parent level. It will be used to generate the hierarchy from which users select object classification information. When the taxonomy structure changes or is added to the master table is updated once. Rather than changing the taxonomy structure on several HTML pages every time there is an update. Ultimately, HTML pages will not be static but generated on the fly using fields extracted from the master tables as required. Field names with the same parent are siblings, or at the same level in the taxonomy.

Forms and browsable examples of the hierarchy are generated when the user selects a particular characteristic by searching the master table. For example, when the provider or consumer selects "Functional Area" all those fields where parent is "Functional Area" are displayed in a browsable manner (like the frames feature in Netscape shown in Figs. 4.1, 4.2 and 4.3). If the user then selects "Legal" from the list of children of "Functional Area" another alphabetical list is created of all of the children of "Legal" and this goes on as deep as the particular branch goes (from one to six levels deep). The consumer defines what he is looking for by selecting the characteristics in a similar manner.

The mechanism by which objects are retrieved from the repository involves searching the repository for an exact match of all fields of the user's search parameters. If all fields are not exactly matched then the system scores those available by relevancy, showing only those that are highly relevant. Relevancy is determined by calculating the distance between the object the user seeks and objects that are registered in the repository. Those elements that belong to the same parent are scored the highest and those that share a parent in common but two levels higher receive a lesser relevant score. This is, of course, going on in each of the six characteristics. These concepts were discussed in Chapter II.

To make the database available with up to the minute accuracy and high relevancy we use batch processing of queries and search using a classification scheme.

B. USER INTERFACE

Once the taxonomies were determined and the system foundation built, the next concern was creating the User Interface to let the user know what the system can do and how to access those capabilities.

1. Literature Review

A literature review of this area revealed many principles on which a system's user interface should be developed. Some of those that were useful included: Donald Norman's book *User Centered System Design: New Perspectives on Human-Computer Interaction* (1986), Ramsey and Atwood's *Human Factors in Computer Systems: A Review of the Literature* (1979), Andriole's book *Handbook of Decision Support Systems* (1989), and Carey's *Human Factors in Management Information Systems* (1988).

2. Principles of User Interface Design

Some of the guiding principles of User Interface taken from these books are:

- Begin by profiling the intended user
- Allow user to enter data by selection not entry
- Predictable behavior to guide the new user
- Use as few keystrokes or steps as possible to perform function (Carey, 1988)

To begin with the first principle of characterizing the user and his skills, Ramsey and Atwood identified three types and associated characteristics:

- Naive user
- Managerial user
- Scientific-Technical user

Some more helpful questions for identifying the user and their use of the system that should influence system design are:

"How frequently will users use the system?

Will the system be used by users under situational pressure?

How experienced with interactive computing are the users?

How experienced with analytical methodology are the users; are they inclined to think analytically or are they more passive users?"(Andriole, 1989)

3. Who is the User?

The range of users anticipated are managers, operations research practitioners, students of OR, and students of decision support and expert systems. Moreover, we want the user interface to be appropriate for a user from DoD or from the general public seeking a decision technology. For this reason we also want the search procedures to be straightforward enough for the typical Internet user. Users of DecisionNet will probably be primarily managerial users. Managerial users are characterized as having: variable information needs, very low tolerance for system "impedance". If dissatisfied with the user interface they will use it less or not at all or have someone else use it to perform the functions they need. (Carey, 1988) The point about variable information needs is a reminder that these managers will probably be well educated but from diverse backgrounds and have vocabulary and definitions of key terms based on their field or background.

There may also be naive users of the system who need an online tutorial (which can be provided through online documentation). Scientific-technical users will have very proprietary definitions of key terms and would benefit from documentation explaining the use of terms and the design of system. In fact, any application should have at a minimum strong feedback mechanisms and built in documentation, regardless of the users' skill level.

We are designing the system based on the following assumptions about the users and their skill levels:

Internet user knows how to point and click and is comfortable with Netscape user interface, drop down select boxes or frames. The user may not know anything about decision support systems except what has been explained on entry pages.

Undergraduate/Graduate student- Information Systems or Operations Research student who is looking for DSS's. This user may need clarification on terms and solution method taxonomy.

Manager (military or civilian) This user is looking for a DSS or expert system for a specific application in their industry or problem area. Solution method probably isn't important but as mentioned earlier this user will have difficulty with distinctions made in the vocabulary.

Academia- Familiar with Internet and decision support systems but may need clarification of terms and definitions. This user is likely to be frustrated if user interface restricts what they want to do and would appreciate the capability to jump between areas they are browsing.

The minimum skill set is relatively high because DSS's are management tools and this is an educated user group. Users may not have a sufficient background in OR to select a solution method from a complex classification of systems. Solution method classification will be of primary interest to academia and is essential to provider registration of objects.

4. Application of User Interface Design Principles

The principles listed previously were specifically applied to DecisionNet through the User Interface design. Users select rather than create items for data entry from dropdown menus for classification of objects. DecisionNet pages and methods for executing functions provide for a consistent user interface and meaningful feedback. These qualities help the user to learn system and how it works. The use of Netscape's Internet browser, forms and dropdown menus minimizes the number of steps the user has to perform. Most are initiated by pointing and clicking by the user. Forms provide the perfect vehicle for capturing information because they are a familiar user interface (people know what is asked of them and how to reply) and they are a logical structure for handling database field names. Also the dropdown menus used with forms minimize the amount of information users have to remember between browsing the taxonomy and stating their search parameters.

The user interface was implemented with some of the tools and capabilities of HTML and Netscape Internet browsers.

C. PRESENTING TAXONOMIES AND CLASSIFICATIONS TO USERS

A challenge for this element of the thesis is to effectively represent the taxonomy to the users. During this thesis work, more and more tools became available for the implementation of a navigatable tree using Netscape browser software. The first tool was "frames" which became widely available when Netscape 2.0 was in testing and released in February 1996. The taxonomy representation structure stored every taxon by its own name and that of its parent, in a table as shown in Tables 4.1 and 4.2. When "Industry" is selected, all of the elements in the Industry master table with "Industry" as a parent are displayed. When "Manufacturing" is selected, all of the elements in the Industry master table with "Manufacturing" as the parent are shown. This was adapted to our relational database design and the taxonomy representation which can be seen in Fig. 4.1, 4.2 and 4.3.

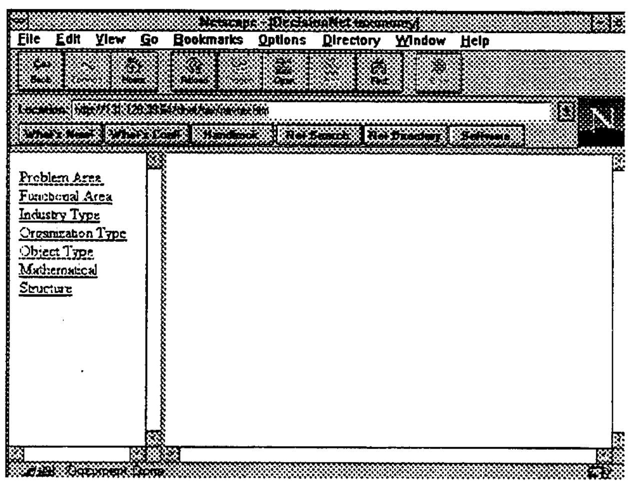

Figure 4.1 All Taxonomies at Level 0

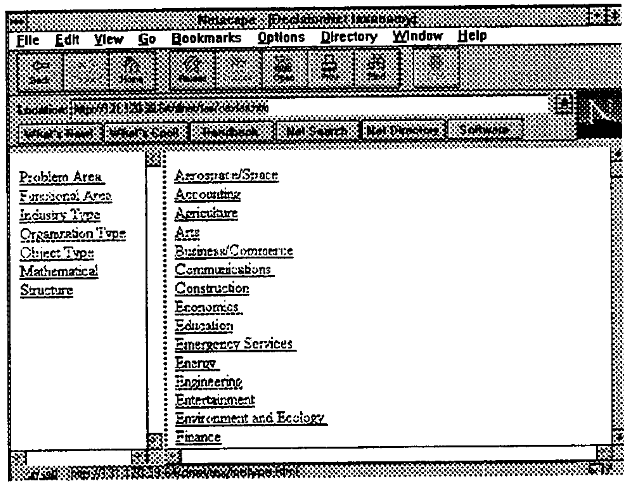

Figure 4.2 Industry Taxonomy at Level 1

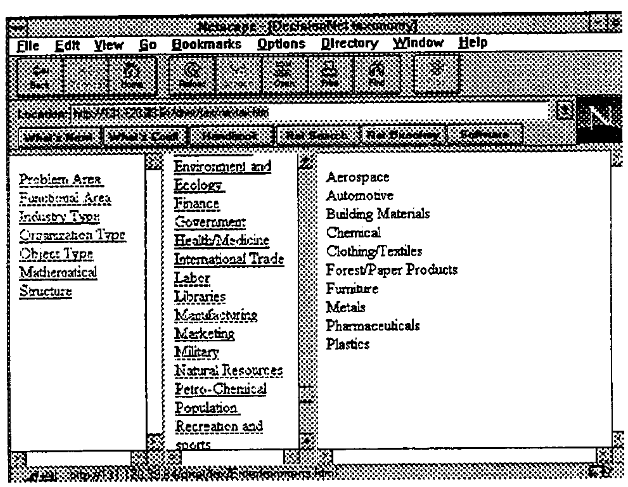

Figure 4.3 Industry/Manufacturing at Level 2

Because the DecisionNet user is remote and navigating the taxonomy to create their search parameters, a difficult aspect of this implementation was keeping memory of past selections between the online forms and taxonomy representation. The database structure (shown in Fig 4.1, 4.2 and 4.3) is fine for browsing the taxonomy, but becomes complicated when a selection made while browsing results in the execution of several scripts and database queries on the DecisionNet system. This is because the transfer of files from server to client using the World Wide Web is a onetime transaction; the packet is sent and the brief connection broken. Each new request generates a new attempt to log into the server and establishes a new connection. Maintaining state between views of the taxonomy is an important part of the user interface because of the complexity of sorting through six taxonomies. Because of the complexity of passing elements of data through CGI and the users unstructured browse technique, the DecisionNet development team is waiting for the arrival of another Delphi companion product called "WebHub" developed specifically for this type of problem. This problem is called "maintaining state" while a user is logged into and browses a site generating multiple queries and scripts to execute. The developer Ann Lynnworth compares it to "keeping track of the items a shopper puts in their grocery cart", (Lynnworth, 1995) which is something current "onetime submit" forms do not do well.

D. GETTING INFORMATION FROM THE USER

The User Interface to DecisionNet is a series of HTML pages that the user navigates beginning with their login to DecisionNet. The initial HTML pages are user or provider registration and user or provider login. These are static HTML pages using forms and table HTML tools. The Delphi script executes and either accepts or rejects the login or registration information. Other functions are offered to users after the login is verified against previously entered login information. Ideally these pages providing other functions would not actually exist on the server but would be created for users to access when a user is correctly logged in (to ensure that security or billing functions are being enforced). This is also because the bulk of the information they contain is generated from the current database structure. For this implementation phase and to get the scripts working I created actual .html files on the server that the Delphi script points to when it has successfully executed. This is done by passing HTML commands to the server browser, which it displays. Because we want the user to select from a list of choices for each classifying characteristic, drop down lists on registration and search forms are used. This eliminates the opportunity to misspell or forget which choice is appropriate while switching from one HTML page to another. An example of a dropdown list on a form is shown in Fig. 4.4.

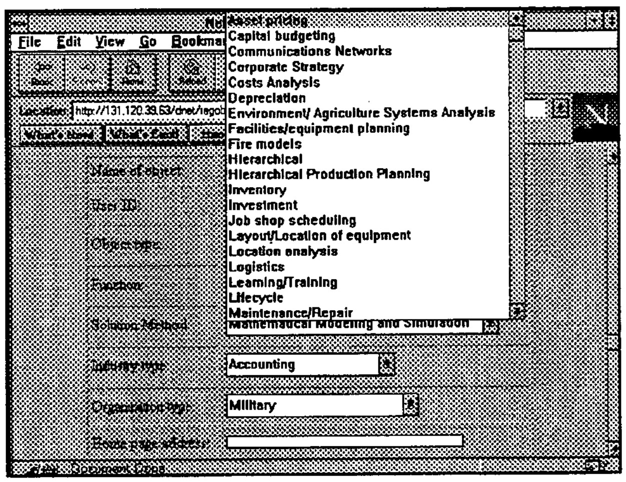

Figure 4.4 Example of Dropdown Menu on the Registration Page

The ideal (though complicated to code) interface would be navigatable frames within a form that build the taxonomies for a search and perform possibly several searches in a row while staying at what is essentially the same first page. Using frames tools a static shell in the margins could be much like the toolbars and powerbars in Windows-based applications. In this example, the content changes but the framework stays the same, which is pleasing as a user interface and provides a consistent location for functions, definitions and online help.

Frames (as shown in Fig. 4.1, 4.2 and 4.3) can be used to implement online documentation which will be helpful to users in deciding what they're looking for. A distinct advantage of using frames is it will allow tutorials to not be disruptive to advanced users. Thoroughness to inexperienced users can be distracting to experienced users. Moreover, the use of frames and HTML pages gives online help for other questions, further explanation, and in the future, online tutorial or help screens through frames. Another possibility for DecisionNet is the opportunity to send the entire taxonomy classification to the client as an executable program within the browser. The user could then navigate the taxonomy and make their selections and send back their search parameters to the DecisionNet system. In this scenario the user queries the DecisionNet server and database once, rather than reconstruct the taxonomy classification and query the database to navigate the taxonomy several times per user session.

HTML tools are available for building an online system that is much more than a terminal linking to a distant computer but a system that provides a clear, meaningful and instructive user interface to decision support tools and helps users find objects they may not have realized were available.

V. CONCLUSIONS

A. SUMMARY

The DecisionNet system aims to provide decision makers and researchers global access, over the Internet, to a large distributed collection of decision technologies. Recent research in OR/MS and Information Technology involves several other efforts with similar objectives. Successful use of such repositories depends, to a large extent, on the users being able to search this large collection to locate relevant technologies. This depends, in turn,on i) a suitable indexing and classification scheme along which the technologies are organized, and ii) intelligent search and retrieval algorithms for computing relevance of technologies against given search criteria.

This thesis has addressed the first part of this problem. Specifically we have developed and constructed taxonomies to allow for indexing and classification of decision technologies in the repository. These technologies include data sets, algorithms, models and decision support systems. Previous research in this area has resulted in the development of few indices of decision technologies; the most notable effort is the Guide to Available Mathematical Software (Boisvert, 1991). Another, Yahoo (Yahoo, 1995), is a thorough subject indexing scheme. However, these index hierarchies are single dimensional, in most cases the dimension is the solution methodology used in the technology.

An early conclusion of our research was that, in order to be effective from a user's point of view, technologies had to be classified along several different dimensions. Five primary dimensions were identified: these are functional area, problem area, industry, organization type and solution method. These dimensions represent the likely starting points of a user's search for a suitable technology and encompass search requirements for a diverse group of users including researchers, analysts, and end users. The detailed taxonomies were developed by combining the terms in existing taxonomies with criteria and principles for taxonomy development. A database architecture was developed to store both the taxonomies (including structural relationships between terms in the taxonomies) and the decision technology objects in the repository. A WWW-capable user interface was also developed to allow providers to register and index their technologies and to allow users to browse through the taxonomies or to specify search criteria.

B. FURTHER RESEARCH

As mentioned above, this thesis addressed one part (development of a classification scheme) of the indexing and retrieval problem for a repository of decision technologies. Further research is needed on the second part, to develop intelligent search and retrieval algorithms. The basic tradeoff in this area is between recall and relevance. An algorithm that insists on a very tight match will probably find highly relevant objects, but it may fail to recall others that are relevant. An algorithm that applies several relaxation rules (e.g., synonyms, generalizations) in conducting the search will recall many more objects, but many of them may not be relevant. The ultimate test of a classification scheme is how useful it is in solving the user's search/retrieval problem. Several algorithms need to be developed and tested in order to determine the usefulness of the classification scheme laid out in this thesis.

Another area of research involves the development of software agents to perform administrative and maintenance tasks on the classification scheme itself. Due to the nature of decision technologies any classification scheme will need to undergo changes over time. How will such changes affect the existing structures and the classifications already made with the existing taxonomies? With good maintenance agents, these changes should be achieved with minimal human intervention.

This thesis has assumed a user interface based on the World Wide Web's client-server model. For example, a provider wishing to index a technology at a lower level in the taxonomy may have to send several messages to the index server. It may be useful to investigate other mechanisms for accomplishing such tasks. For example, the Java language may be used to create an applet that is transferred to the provider's machine and which allows the provider to browse through the taxonomy levels without having to send successive requests to the index server.

C. CONCLUSIONS

The main contribution of this thesis is in creating taxonomies that can be used to classify decision support systems. This taxonomy development has involved the consideration of the viewpoints of several different kinds of users, issues in user interface design, as well as a systematic application of principles of taxonomy development. While this work has been done in the context of decision technologies, some of the ideas presented here may be generalized to other kinds of information and physical products offered on the Internet.

The increase in the amount of information and in the number of information-publishers, and the exponential growth of Internet users, makes the indexing and retrieval problem more critical. The practice of Information Brokering that charges users for accurate, verified and relevant information and for search tools will undoubtedly increase. A strong and robust classification system is vital to the effort to organize and search information about a collection of objects.

LIST OF REFERENCES

Andriole, Stephen J., *Handbook of Decision Support Systems*, Tab Books, Blue Ridge Summit, PA, 1989.

Bhargava, Hemant K., A.S. King and D.S. McQuay, "DecisionNet: An Architecture for Modeling and Decision Support Over the World Wide Web," *International Symposium on Decision Support Systems*, Hong Kong, July 1995.

Bosivert, Ronald, "The Architecture of an Intelligent Virtual Mathematical Software Repository System," *Mathematics and Computers in Simulation,* 36 Elsevier Science Netherlands, 1994.

Boisvert, Ronald, "The Guide to Available Mathematical Software Problem Classification System," *Communications in Statistics, part B-Simulation and Computation,* 20(4), 1991.

Boisvert, Ronald, "Preliminary Design of a Taxonomy for Mathematical Software," Working Note, August 1994.

C. Mic Bowman, Peter B. Danzig, Darren R. Hardy, U. Manber, and M.F. Schwartz. "The Harvest Information Discovery and Access System," Proceedings of Second International World Wide Web Conference, October 1994.

Calvert, Charles. *Delphi Unleashed,* Sams Publishing, Indianapolis, IN, 1995.

Carey, Jane M., ed., *Human Factors in Management Information Systems*, ABLEX Publishing Corp, Norwood New Jersey, 1988.

Defense Information Systems Agency, "Global Command and Control System Common Operating Environment Requirements," DISA, August 1994.

Digital Corporation, "Digital's Industry Taxonomy",Digital, *VAX VMS Software Source book* (7th ed.), Digital Corporation, 1991.

Gass, Saul I. and Carl M. Harris, eds., *Encyclopedia of Operations Research and Management Science*, Kluwer Academic Publisher, 1995.

Glass, Robert and Iris Vessey, "Towards Taxonomies of Application Domains: Contemporary Candidates", 1994.

IBM Corporation, *Industry Applications and Abstracts* (10th ed.) IBM Corp. 1988.

King, Andrew, LT USN, "DecisionNet -- A Prototype of Distributed Decision Support Systems," Naval Postgraduate School Thesis, September 1995.

Lynnworth, Ann, HREF Tools, Corp., (http://www.href.com), 1995.

McQuay, Daniel, CAPT USMC, "A Framework for a Distributed Decision Support Network," Naval Postgraduate School Thesis, September 1995.

Nagel, Stuart S., ed., *Computer Aided Decision Analysis: Theory and Applications,* Westport, Conn, Quorum Books, 1993.

Norman, Donald A. and Stephen W. Draper, *User Centered System Design: New Perspectives on Human-Computer Interaction,* Lawrence Erlbaum Associates, Hillsdale, NY, 1986.

North Carolina State University Water Quality Group, "Water, Soil, Hydro-Environmental Decision Support System (WATERSHEDSS)," (http://h2osparc.wq.ncsu.edu/), August, 1994.

Ramsey, R. and M. Atwood, *Human Factors in computer systems: A Review of the Literature*, Science Applications, 13, 1979.

Reifer Consultants, "Reifer's Application Taxonomy," Productivity and Quality Survey, 1990.

Sprague, Ralph, "A Framework for the Development of Decision Support Systems," *MIS Quarterly*, 4(2), June 1980.

Sun Microsystems, Inc., "About Java", (http://www.javasoft.com/about.html), 1995.

Tolle, Jon W., ed., *OR/MS Index 1982-1987*, Institute of Management Sciences, Providence, RI and Operations Research Society of America, Baltimore, MD, 1988.

Wall, Larry and Randal L. Schwartz, *Programming Perl*, O'Reilly &Associates, Inc., 1991.

Yahoo! Corp., (http://www.yahoo.com), 1995.

APPENDIX: SOLUTION METHOD TAXONOMY

A Simulation/ stochastic modeling
- A1. Simulation
- A1a. Discrete
- A1b. Continuous (Markov models)
- A2. Queueing
- A3. Reliability
- A3a. Quality control
- A3b. Electrical network

B Statistics/probability - problem has elements of uncertainty
- B1. Aproximation
- B1a. Least squares (L_2) approximation
- B1a1. Linear least squares
- B1a2. Nonlinear least squares
- B1b. Minimax (L_infinity) approximation
- B1c. Least absolute value (L_1) approximation
- B1d. Other analytic approximations (e.g., Taylor polynomial, Pade)
- B1e. Smoothing
- B1f. Service routines for approximation
- B1f1. Evaluation of fitted functions, including quadrature
- B1g. Grid or knot generation
- B1h. Manipulation of basis functions (e.g., evaluation, change of basis)
- B1i. Other
- B2. Statistics, Probability
- B2a. Data summarization
- B2a1. One-dimensional data
- B2a2. Two dimensional data
- B2a3. Multi-dimensional data
- B2b. Data manipulation
- B2b1. Transform (search also classes L10a1, N6, and N8)
- B2b2. Tally
- B2b3. Subset
- B2b4. Merge
- B2b5. Construct new variables (e.g., indicator variables)
- B2c. Elementary statistical graphics
- B2c1. One-dimensional data
- B2c2. Two-dimensional data
- B2c3. Three-dimensional data
- B2c4. Multi-dimensional data
- B2d. Elementary data analysis

B2d1. One-dimensional data
B2d2. Two-dimensional data
B2d3. Multi-dimensional data
B2d5. Multiple Multi-dimensional data sets
B2e. Function evaluation
B2e1. Univariate
B2e2. Multivariate
B2f. Random number generation
B2f1. Univariate
B2f2. Multivariate
B2f3. Service routines (e.g., seed)
B2g. Analysis of variance (including analysis of covariance)
B2g1. One-way
B2g2. Two-way
B2g3. Three-way (e.g., Latin squares)
B2g4. Multi-way
B2g5. Multivariate
B2g6. Generate experimental designs
B2g7. Service routines
B2h. Regression
B2h1. Simple linear (i.e., y = b_0 + b_1x)
B2h2. Polynomial (e.g., y = b_0 + b_1x + b_2 x^2)
B2h3. Multiple linear (i.e., y = b_0 + b_1 x_1 + ... + b_p x_p)
B2h4. Polynomial in several variables
B2h5. Nonlinear (i.e., y = F(X,b))
B2h6. Simultaneous (i.e., Y = Xb)
B2h7. Spline (i.e., piecewise polynomial)
B2h8. EDA (e.g., smoothing)
B2h9. Service routines (e.g., matrix manipulation for variable selection)
B2i. Categorical data analysis
B2i1. 2-by-2 tables
B2i2. Two-way tables
B2i3. Log-linear model
B2i4. EDA (e.g., median polish)
B2j. Time series analysis
B2j1. Univariate
B2j2. Two time series
B2k. Correlation analysis (search also classes L4 and L13c)
B2l. Discriminant analysis
B2m. Covariance structure models
B2n. Cluster analysis

B2n1. One-way
B2n2. Two-way
B2n3. Display
B2n4. Service routines (e.g., compute distance matrix)
B2o. Life testing, survival analysis
B2p. Multidimensional scaling
B2q. Statistical data sets

C Gaming - models adversarial conflict between intelligent opponents /Game theory

D Optimization and Linear Algebra

D1. Linear Algebra
D1a1. Elementary vector and matrix operations
D1a2. Elementary matrix operations
D1b. Solution of systems of linear equations (including inversion, LU and related decompositions)
D1b1. Real nonsymmetric matrices
D1b2. Real symmetric matrices
D1b3. Complex non-Hermitian matrices
D1b4. Complex Hermitian matrices
D1b5. Associated operations (e.g., matrix reorderings)
D1c. Determinants
D1c1. Real nonsymmetric matrices
D1c2. Real symmetric matrices
D1c3. Complex non-Hermitian matrices
D1c4. Complex Hermitian matrices
D1d. Eigenvalues, eigenvectors
D1d1. Ordinary eigenvalue problems (Ax = lambda x)
D1d2. Generalized eigenvalue problems (e.g., Ax = lambda Bx)
D1d3. Associated operations
D1e. QR decomposition, Gram-Schmidt orthogonalization
D1f. Singular value decomposition
D1g. Update matrix decompositions
D1g1. LU
D1g2. Cholesky
D1g3. QR
D1g4. Singular value
D1h. Other matrix equations (e.g., AX+XB=C)
D1i. Singular, overdetermined or underdetermined systems of linear equations, generalized inverses
D1i1. Unconstrained
D1i2. Constrained
D1i3. Generalized inverses

D2 Optimization
D2a. Unconstrained
D2a1. Univariate
D2a2. Multivariate
D2b. Constrained
D2b1. Linear programming
D2b2. Transportation and assignments problem
D2b3. Integer programming
D2b3a. Zero/one
D2b3b. Covering and packing problems
D2b3c. Knapsack problems
D2b3d. Matching problems
D2b3e. Routing, scheduling, location problems
D2b3f. Pure integer programming
D2b3g. Mixed integer programming
D2b4. Network (for network reliability search class M)
D2b4a. Shortest path
D2b4b. Minimum spanning tree
D2b4c. Maximum flow
D2b4d. Test problem generation
D2b5 Quadratic programming
D2b6. Geometric programming
D2b7. Dynamic programming
D2b8. General nonlinear programming
D2b9. Global solution to nonconvex problems
D2c. Optimal control
D2d. Service routines
D2d1. Problem input (e.g., matrix generation)
D2d2. Problem scaling
D2d3. Check user-supplied derivatives
D2d4. Find feasible point
D2d5 Check for redundancy
D2d6. Other

E Numerical Math -(non optimization)/(non statistics)
E1. Arithmetic, error analysis
E1a. Integer
E1b. Rational
E1c. Real
E1c1. Standard precision
E1c4. Extended precision
E1c4. Extended range
E1d. Complex

E1d1. Standard precision
E1d3. Extended precision
E1d4. Extended range
E1e. Interval
E1f. Change of representation
E1f1. Type conversion
E1f2. Base conversion
E1f3. Decomposition, construction
E1g. Sequences (e.g., convergence acceleration)
E2. Number theory
E3. Elementary and special functions
E3a. Integer-valued functions (e.g., factorial, binomial coefficient, permutations, combinations, floor, ceiling)
E3b. Powers, roots, reciprocals
E3c. Polynomials
E3c1. Orthogonal
E3c2. Non-orthogonal
E3d. Elementary transcendental functions
E3d1. Trigonometric, inverse trigonometric
E3d2. Exponential, logarithmic
E3d3. Hyperbolic, inverse hyperbolic
E3d4. Integrals of elementary transcendental functions
E3e. Exponential and logarithmic integrals
E3f. Cosine and sine integrals
E3g. Gamma
E3g1. Gamma, log gamma, reciprocal gamma
E3g2 . Beta, log beta
E3g3. Psi function
E3g4 . Polygamma function
E3g5. Incomplete gamma
E3g6. Incomplete beta
E3g7. Riemann zeta
E3h. Error functions
E3h1. Error functions, their inverses, integrals, including the normal distribution function
E3h2. Fresnel integrals
E3h3. Dawson's integral
E3i. Legendre functions
E3j. Bessel functions
E3j1. J, Y, H_1, H_2
E3j2. I, K
E3j3. Kelvin functions

E3j4. Airy and Scorer functions
E3j5. Struve, Anger, and Weber functions
E3j6. Integrals of Bessel functions
E3k. Confluent hypergeometric functions
E3l. Coulomb wave functions
E3m. Jacobian elliptic functions, theta functions
E3n. Elliptic integrals
E3o. Weierstrass elliptic functions
E3p. Parabolic cylinder functions
E3q. Mathieu functions
E3r. Spheroidal wave functions
E3s. Other special functions
E4. Interpolation
E4a. Univariate data (curve fitting)
E4a1. Polynomial splines (piecewise polynomials)
E4a2. Polynomials
E4a3. Other functions (e.g., rational, trigonometric)
E4b. Multivariate data (surface fitting)
E4b1. Gridded
E4b2. Scattered
E4c. Service routines for interpolation
E4c1. Evaluation of fitted functions, including quadrature
E4c2. Grid or knot generation
E4c3. Manipulation of basis functions (e.g., evaluation, change of basis)
E4c4. Other
E5. Solution of nonlinear equations
E5a. Single equation
E5a1. Polynomial
E5a2. Nonpolynomial
E5b. System of equations
E5c. Service routines (e.g., check user-supplied derivatives)
E6. Aproximation
E6a. Least squares (L_2) approximation
E6a1. Linear least squares
E6a2. Nonlinear least squares
E6b. Minimax (L_infinity) approximation
E6c. Least absolute value (L_1) approximation
E6d. Other analytic approximations (e.g., Taylor polynomial, Pade)
E6e. Smoothing
E6f. Service routines for approximation
E6f1. Evaluation of fitted functions, including quadrature

E6g. Grid or knot generation
E6h. Manipulation of basis functions (e.g., evaluation, change of basis)
E6i. Other
E6a1. Unconstrained
E6a2. Constrained
E6b. Nonlinear least squares
E6b1. Unconstrained
E6b2. Constrained

F Spreadsheet modeling (not include imbedded optimization) formula/expression evaluation

G Knowledge based

H Symbolic math (calculus)

H1. Elementary and Special Functions
H1a. Integer-valued functions (e.g., factorial, binomial coefficient, permutations, combinations, floor, ceiling)
H1b. Powers, roots, reciprocals
H1c. Polynomials
H1c1. Orthogonal
H1c2. Non-orthogonal available only on grid
H2. Differnetiation, integration
H2a.
H2b. Multidimensional integrals
H2b1. One or more hyper-rectangular regions (includes iterated integrals)
H2b2. n-dimensional quadrature on a nonrectangular region
H2c. Service routines (e.g., compute weights and nodes for quadrature formulas)
H3. Differential and integral equations
H3a. Ordinary differential equations (ODE's)
H3a1. Initial value problems
H3a2. Multipoint boundary value problems
H3a3. Service routines (e.g., interpolation of solutions, error handling, test programs)
H3b. Partial differential equations
H3b1. Initial boundary value problems
H3b2. Elliptic boundary value problems
H3c. Integral equations
H4. Integral transforms
H4a. Trigonometric transforms including fast Fourier transforms
H4a1. One-dimensional
H4a2. Multidimensional
H4b. Convolutions

H4c. Laplace transforms
H4d. Hilbert transforms

I Decision Analysis

I1. Decision Trees
I2. Utility Theory
I3. Influence Diagrams
I4. AHP
I5. Risk Analysis

J Other

INITIAL DISTRIBUTION LIST

1.	Defense Technical Information Center 8725 John J. Kingman Rd. STE 0944 Ft. Belvoir, VA 22060-6218	2
2.	Dudley Knox Library Naval Postgraduate School 411 Dyer Rd. Monterey, CA 93943-5101	2
3.	Professor Hemant Bhargava (Code SM/BH) Naval Postgraduate School Systems Management Department Monterey, CA 93943-5000	3
4.	Professor Gordon Bradley (Code OR/BZ) Naval Postgraduate School Operations Research Department Monterey, CA 93943-5000	2
5.	CAPT George Zolla Naval Postgraduate School Monterey, CA 93943-5000	1
6.	LT Patricia Rogers 9981 Waldgrove Pl. San Diego, CA 92131	2